

INTRODUCTION TO THE MM5 3D-VAR DATA ASSIMILATION SYSTEM: THEORETICAL BASIS

F. Vandenberghe and Y.-H. Kuo

*National Center for Atmospheric Research
P.O. Box 3000
Boulder, CO 80307-3000*

1. INTRODUCTION

Data assimilation has long been regarded primarily as a mean of providing initial conditions for Numerical Weather Prediction (**NWP**) in meteorological centers. Increasingly, it is now being recognized that, through the constant confrontation of theory (under the form of a numerical model that discretized the physical laws governing the atmospheric flow) with reality (as depicted by meteorological observations), the data assimilation process has the potential to bring major advances in our scientific understanding of the atmosphere. An ideal assimilation should be able to process all the available information (*i.e.* the observations themselves, the meteorological model, and the known statistical properties of the flow) together with the uncertainties of these various sources of information to produce a complete and consistent description of the flow with its associated uncertainty. Data assimilation is, an estimation problem and Estimation theory constitutes the natural mathematical foundation for understanding data assimilation problems. It provides, in addition, a number of algorithms for approaching and addressing these problems. However, as it was pointed out by Cohn (1997), while engineering problems, for which Estimation Theory has primarily been developed, are generally small-scale and sometimes linear, atmospheric data assimilation problems are large-scale and generally non-linear. Sensible computational approximations have, therefore, to be made to implement these algorithms in atmospheric and oceanic application. Indeed, most of the research work performed on data assimilation is intended at determining cost effective simplifications to implement.

From this standpoint, the introduction in the 80's of the Variational Calculus in the data assimilation field represents certainly one of the most important breakthrough of these researches. The variational formalism is used to reformulate the data assimilation problem in term of optimization and it offers an efficient way to perform this optimization. Applied

either as the so-called 3-Dimensional Physical Space Analysis System (**3D-PSAS**) or the so-called 3-Dimensional Variational (**3D-VAR**) data assimilation system, the variational calculus has opened new horizons in the atmospheric and oceanographic data assimilation field. Recently, using a 3D-PSAS approach, Cohn *and al.* (1998) were able to solve the classical Optimal Interpolation problem without the traditional data selection procedure; an approximation that has been in force in meteorological centers for decades (Lorenç 1986).

Moreover, the 3D-PSAS and 3D-VAR formalism naturally generalizes in the time dimension into, respectively, 4D-PSAS and 4D-VAR data assimilation. Lorenç (1986) has shown that under the perfect linear model assumption and in presence of white Gaussian noise, 4D-PSAS and 4D-VAR produce the same result at the end of the assimilation period as a Kalman Filter (and as a Kalman smoother over the whole assimilation interval). The Kalman filter and smoother have the very interesting property to be "optimal", in the sense that the Kalman filter minimizes the estimation error at the time of the analysis while the Kalman smoother minimized this error over the whole assimilation interval. In addition, Kalman algorithms provide an explicit way to compute forecast errors. Although, standard Kalman algorithms can only be considered as prototype algorithms for data assimilation in view of the many assumption involved and of their computational requirements; they still provide theoretical references and useful interpretations in term of information and uncertainty for the design of practical data assimilation systems. Because of these interesting features, there have been tremendous efforts in the atmospheric community and particularly in operational centers to develop new forecasting systems based on variational algorithms. So far, the 3D/4D-VAR approach has been the most popular choice. The reasons for this preference are, however, mostly historical since, in essence, 3/4D-PSAS 3/4D-VAR are completely equivalent data assimilation systems in term of both results and overall cost (Courtier 1997). At the present time, 3D-VAR data assimilation systems are already operational at NCEP (Parrish and Derber 1992), ECMWF (Courtier *et al.* 1997), RPN in Canada, the UK Met Office and Météo France, while the NASA Data Assimilation Office is routinely running a 3D-PSAS data assimilation procedure.

The necessary components for a 3-dimensional data assimilation system are a forecasting model, an observing system and the statistics on the accuracy of the forecasts and measurements. At the mesoscale ($1 \sim 100\text{km}$), these statistics, and, in particular the

information related to forecast errors, are certainly the least known and the most difficult to obtain of the three components. Indeed, if operational centers have a long experience and extensive forecast data bases from which statistics can be deduced and studied (Hollingsworth and Loennberg 1986, Loennberg and Hollingsworth 1986, Philipps 1986), such experience does not exist for mesoscale models. There is, therefore, an important research work to be accomplished in this domain and these notes describe the early results obtained in that domain with the Penn State/NCAR Mesoscale Model version 5, **MM5** hereafter. MM5 is a non-hydrostatic primitive equation model (Dudhia 1993), for which the adjoint operator has been developed (Zou *et al.* 1997) and have been used for data assimilation purposes (Zou and Kuo 1996) with basic statistical assumption (smoothing). These studies have evidenced the limitation of these assumptions and called for a more realistic statistical model. Our researches has focused on the the so-called *NMC method* (Parrish and Derber 1992, Rabier *et al.* 1998) and have leaded into the development of a mesoscale 3D-VAR system, the natural testbed to conduct such researches. Practical information on the use of this system will be published apart, the goal of the present notes is rather to give an general introduction to 3D-VAR data assimilation techniques and the specific issues related to mesoscale applications.

Because, statistical aspects are of primary concern, these notes have been organized around this central question. Having presented the 3-dimensional data assimilation problem in section 2, the basic elements of Estimation Theory are then introduced and developed all over section 3. In particular, we pose the atmospheric data assimilation problem as an estimation problem and we show that, using the probabilistic approach of the Estimation theory framework, it is possible to quantify the concepts of information and uncertainty. As a result, the solution of the 3-dimensional data assimilation problem can be found by minimazing the uncertainty present in the system. An algorithm to compute the solution corresponding to that minimum is proposed in section 4. Details on practical implementation are discussed in section 5 and some numerical results are presented in section 6. Concluding remarks are gathered in section 7.

2. THREE – DIMENSIONAL DATA ASSIMILATION

The aim of meteorological data assimilation is to estimate the evolution of the atmosphere as accurately as possible. For that purpose, meteorologists make use of NWP models. Computer-based NWP models are mathematical representations of the physics governing atmospheric motions. Through the solving of the model equations, computers can yield quantitative estimates of the evolution of the atmosphere and thus provide weather forecasts. Limitations, however, are inherent in atmospheric models. One of these is our imperfect knowledge of meteorological phenomena; another is the fact that our representation of the continuous actual physical systems must be discrete within the models. Thus, models can only approximate the real atmosphere and true conditions. Without observational information to help correct the model representation of the atmosphere, predictions would rapidly deviate from reality, with large forecast errors prevailing. To maximize forecast accuracy, therefore, observational data must be used to correct for model correction.

The 3-dimensional solution of the data assimilation problem consists in regular re-initializations of the forecasting model with initial conditions derived from the latest observations that have been made available. Thus, 3-dimensional data assimilation is a succession of cycles made of initialization and forecast operations. For a given model, the forecast operation consists in an integration and is relatively straitforward, although generally the most time consuming. The initialization procedure is much more complex. The difficulty arises from the facts that observations come with different nature, different distribution and different accuracy. They must, therefore, be processed or, at least, analysed prior to be used. Moreover, all the observations valid at a single time are far insufficient to give alone a complete description of the atmosphere and an additional information is required to supplement the missing information. This additional information, referred to as background information, is usually provided by NWP models under the form of forecast issued from a previous analysis.

Hence, meteorological analysis attempts to combine several sources of data, each source bringing simultaneously its own content of information and uncertainty that needs to be quantified. It is a classical results of information theory (Shannon 1948, Haming 1988) that uncertainty is measured by means of probability density function and it is the

purpose of Estimation theory to minimize uncertainty. Thus, Estimation theory offers the natural framework to tackle the data assimilation problem. This framework is introduced in the next section.

3. ESTIMATION THEORY FOR METEOROLOGICAL ANALYSIS

3.1 Estimation theory framework

3.1.1 Stochastic modeling

Imperfect NWP models and inaccurate measurements are primarily responsible for the presence of uncertainty in the data assimilation system. During the analysis process, information undergoes several transformations that may alter the total amount of uncertainty content in the system. At the end of the process, the remaining uncertainty will be responsible for analysis errors. Clearly, we are interested in data assimilation systems that reduce and not amplify the uncertainty initially present in the input data. In order to find such systems, we need a quantitative way to measure and track uncertainty in data assimilation systems. The best way to mathematically handle uncertainty is probably offered by Stochastic Calculus. The idea of the stochastic approach is that, if it is impossible to predict exactly the unfolding of a phenomenon, we can, at least, apprehend the phenomenon in its generality through its statistical properties. Simple stochastic representations generally include a determinist mean term plus a perturbation that randomly fluctuate from a certain amplitude around zero. Thus, at a given time, the exact state of the atmosphere depends on the particular realization of the random fluctuation and is, therefore, unpredictable from a classical determinist viewpoint. Its behavior can, however, be described and, to some extent, be predicted using a probabilistic approach. In such approach, the atmosphere is described by a probability density function, **pdf** hereafter, which quantify the likelihood for a particular event to occur. It's note the scope of these notes to present probabilities in details, Readers are referred to Jazwinsky (1971) or Roberts (1992) to find more about pdf and stochastic calculus. In the following, we present a few concepts that will be needed to introduce the framework of atmospheric data assimilation.

3.1.2 Estimator and estimate

Practically, the estimation procedure will consist in finding a mathematical function, called the estimation function, of the the background and observation variables that will

reduce and preferably minimize the errors that result from uncertainty present in the input data.

Note that, since the input variables of the estimation function are random its output will be random as well. This variable is called the estimator. The estimator is stochastic. This is a purely conceptual quantity that depends on observing systems only. The estimator should be distinguished from its determinist realization, called the *estimate* which depends on real measurements, those produced by the observing systems described in the estimator. The estimate is the final results of the analysis and ultimately the quantity useful for practical application.

3.1.4 Optimal estimation

The first property we would ask to an estimate is to be unbiased, which means that there is no systematic error. Thus, if the estimation process is repeated many times or indefinitely, we would find in average the true state the atmodphere. We would also like the estimation error to be minimal. This suposes, of course, that we are able to mesure this error. Estimation Theory defines the estimation error as the statistical *variance* of the estimator which, in practice, is given by the trace of the error covariance matrix (Jazwinsky 1971). Based on this measure, the theory shows that the *conditional mean* estimate is always unbiased and minimizes the variance independently of the nature of the models and the probability densities of the various input data (Cohn 1997). For that reason, the conditional mode estimate is said to be optimal and should be sought whenever is possible,

However, this is not always possible in practice. Here, the name "conditional mean" stands for "the mean of the conditional a posteriori probability density". The conditional mean estimate requires, therefore, the evaluation of an integral of the conditional a posteriori probability density function which is a random function. While, it is relatively straitforward to derive analytically a closed form of this integral in a linear and Gaussian context (Cohn 1997), this becomes rapidly impossible as soon as models become non-linear or input probability densities are non-Gaussian. A numerical evaluation based on Monte-Carlo methods is also not possible because of the dimension of the function to integrate. The conditional a posteriori probability density is a n -dimensional function, where n is the number of analysed variables times the number of points of the analysis grid, which is

typically 10^5 in real atmospheric applications.

Because of the difficulty to obtain the conditional mean estimate, the *conditional mode* estimate is generally preferred. The name "conditional mode" stands for "the mode of the a posteriori conditional probability density". From a computational viewpoint, the evaluation of the conditional mode relies on the search of extrema, which can be efficiently numerically performed using standard optimization techniques. The conditional mode estimate has, in addition, the appealing property to be also the conditional mean when models are linear and input probability densities are Gaussian. This property results from the symmetric unimodal form of the Gaussian distribution for which the mode is equal to the mean. As results the conditional mode estimate is also the best linear unbiased estimate (**BLUE**). That is to say, no other linear estimate performs better then the conditional mode estimate. In conclusion, although the conditional mode is not optimal in general, it has some appealing features that make this form of estimation popular in atmospheric science. The rest of these notes is dedicated to the study and implementation of the conditinal mode estimate. For convenience, from now we will call this estimate **MAP** for maximum *a posteriori* estimate.

In order to derive the expression of the MAP, a certain number of assumptions have to be made. These hypotheses are introduced in the next section and the derivation of the MAP follows.

3.2 State vector representation, error sources and input information

3.2.1 State vector and error representations

Following Cohn (1997), we define the true state of the atmosphere to be estimated as a n -dimensional vector x^t . This vector is generally made of a collection of numerical values taken by a set of atmospheric parameters such as pressure, temperature, wind or humidity at the nodes of a 3-dimensional mesh discretizing the geographical domain under study.

3.2.2 Error sources

Obviosuly, a discrete mathematical representation of the continuum of the atmosphere is somewhat arbitrary and undoubtedly erroneous. Such representation is dictated by the need to fit on our computers and we have, unfortunately, no other choice. Consequently,

a certain representativeness error will be accounted for in the error budget of data assimilation systems. Similarly computer-based NWP models that are used to predict the time evolution of the atmosphere are only a finite and discrete form of the continuum dynamics of the fluid mechanics, this results in a model error that accumulates in time. Observations are a third source of uncertainty. Since no measurement devices are perfect, there will always be a certain measurement error in the data ingested by the data assimilation system.

Since representativeness error and model error depend on the true state of the atmosphere which is unknown, these errors are completely unknown and unknowable from a determinist viewpoint. At the contrary, measurement errors depend only on the technology of the instruments and, thus, could be known in theory. However, there will always be some unpredictable noise (calibration, orbit derive, etc.) that will render a certain amount of measurement error completely unpredictable. For these reasons it is appropriate, and in any case one has little choice other than, to represent representativeness, model and measurement errors in a stochastic way.

More about stochastic representation of errors in data assimilation systems can be found in Cohn (1997). In the following we will directly apply the author's results and assume that the combined effects of the representativeness and measurement errors can be represented by single random perturbation called observational error. This process is assumed to be white *i.e.* uncorrelated in time, and Gaussian *i.e.* with Gaussian pdf, of known mean and covariance matrix. The same assumptions are made for the model error.

3.2.3 Model information

Computer-based NWP models are used to propagate in time the state vector. These models are usually obtained by discretization using numerical techniques such as finite differences, finite elements or spectral methods of the partial differential equations that are assumed to govern the flow of the atmosphere. Numeric Models are convenient media to encapsulate all the statistical, dynamical and physical knowledge that we can *a priori* have on the atmosphere. NWP models are supposed to reflect the evolution of the true state of the atmosphere with a certain degree of uncertainty. They are usually represented by a recurrent relation between two successive states of the atmosphere:

$$x^t(t_{i+1}) = M_i[x^t(t_i)] + \eta_i \tag{3.1}$$

Where $x^t(t_i)$ and $x^t(t_{i+1})$ are, respectively, the true state of the atmosphere at time t_i and t_{i+1} ; η_i is a n -dimensional random vector accounting for the model error at time t_i . η_i is assumed to be white and Gaussian with zero mean and known covariance matrix:

$$E\{ \eta_i \} = 0 \quad (3.2)$$

$$E\{ \eta_i \eta_i^T \} = \mathbf{Q}_i \quad (3.3)$$

3.2.4 Non – linearities

NWP models are non-linear, however, most of the results of the Estimation theory that are used in data assimilation apply for linear model only. As a result, it will often be necessary to linearize the model along its reference trajectory. Thus, we define the so-called tangent linear model:

$$\mathbf{M} = \frac{\partial M}{\partial x} \quad (3.4)$$

And we will denote by M_{t_0, t_i} and \mathbf{M}_{t_0, t_i} the operators which integrate, respectively, the NWP model and its tangent linear approximation from time t_0 to time t_i (resolvent matrix).

3.2.5 Background information

Hypotheses:

Let's assume for the moment that at a given time, let's say t_0 , an unbiased estimator x_0 of the true state of the atmosphere is available and we assume that this estimator has a Gaussian probability density, with known error covariance matrix \mathbf{P}_0 . Hence we have:

$$E\{ x_0 - x^t(t_0) \} = 0. \quad (3.5)$$

$$E\{ (x_0 - x^t(t_0))^T (x_0 - x^t(t_0)) \} = \mathbf{P}_0 \quad (3.6)$$

In addition we assume that this estimator is optimal, *i.e.* it minimizes the estimation error. In practice, the optimality is characterized by the trace of the error covariance matrix \mathbf{P}_0 ; for any other unbiased estimator with error covariance matrix \mathbf{P}'_0 , we will have $trace(\mathbf{P}'_0) \geq trace(\mathbf{P}_0)$.

The Predictionp problem:

Now, the prediction problem is to find an estimator, possibly optimal, of the true state of the atmosphere $x^t(t_i)$ for subsequent times t_i , $i = 1, \dots$. Let's consider the problem at time t_1 , we can use the model information and compute the following estimate:

$$x^f(t_1) = M_{t_0, t_1}[x_0] \quad (3.7)$$

Let's examine the properties of this estimator.

The bias is given by:

$$\begin{aligned} E\{ x^f(t_1) - x^t(t_1) \} &= E\{ M_{t_0, t_1}[x_0] - M_{t_0, t_1}[x^t(t_0)] - \eta_1 \} \\ &= E\{ M_{t_0, t_1}[x_0] - M_{t_0, t_1}[x^t(t_0)] \} \quad \text{since } E\{ \eta_1 \} = 0 \\ &\simeq E\{ \mathbf{M}_{t_0, t_1}(x^f(t_0) - x^t(t_0)) \} \\ &= \mathbf{M}_{t_0, t_1} E\{ x(t_0) - x^t(t_0) \} \\ &= \mathbf{M}_{t_0, t_1} \mathbf{0} \quad \text{by application of (3.5)} \\ &= 0 \end{aligned} \quad (3.8)$$

Thus, if the model is linear or weakly non-linear, the estimator defined by (3.7) is unbiased.

The error covariance matrix is given by:

$$\begin{aligned} \mathbf{P}_1 &= E\{ (x^f(t_1) - x^t(t_1))(x^f(t_1) - x^t(t_1))^T \} \\ &= E\{ (M_{t_0, t_1}[x_0] - M_{t_0, t_1}[x^t(t_0)] - \eta_1) (M_{t_0, t_1}[x_0] - M_{t_0, t_1}[x^t(t_0)] - \eta_1)^T \} \\ &= E\{ (\mathbf{M}_{t_0, t_1}(x_0 - x^t(t_0)) - \eta_1) (\mathbf{M}_{t_0, t_1}(x_0 - x^t(t_0)) - \eta_1)^T \} \\ &= \mathbf{M}_{t_0, t_1} E\{ (x_0 - x^t(t_0))(x_0 - x^t(t_0))^T \} \mathbf{M}_{t_0, t_1}^T + E\{ \eta_1 \eta_1^T \} \\ &= \mathbf{M}_{t_0, t_1} \mathbf{P}_0 \mathbf{M}_{t_0, t_1}^T + \mathbf{Q}_1 \end{aligned} \quad (3.9)$$

Where we have assumed that equation (3.8) hold and made the hypothesis that the model error is not correlated with the intial estimation error $E\{(x_0 - x^t(t_0))\eta_1^T\} = 0$.

If the model is linear, then equation (3.8) and (3.9) are exact and it can shown that the trace of the covariance matrix \mathbf{P}_1 is minimal, *i.e* the estimator defined at time t_1 by (3.7) is optimal (see Gelb 1991 for a demonstration of this result). By iteration, it's easy to show that at any subsequent time t_i , $i = 1, \dots$, the estimator defined by:

$$x^f(t_i) = M_{t_0, t_i}[x_0] \quad (3.10)$$

is optimal. In addition, the pdf of such estimator is Gaussian, (Linear transformations conserve the Gaussian nature of a pdf). Thus, this pdf is completely defined by its mean:

$$E\{ x^f(t_i) \} = E\{ M_{t_0,t_i}[x_0] \} \simeq M_{t_0,t_i}[E\{ x_0 \}] \quad (3.11)$$

and covariance matrix \mathbf{P}_i given by (3.9).

In conclusion, the hypotheses made at the beginning of this section ensure the existence at any time t_i of an optimal estimate (3.10) with Gaussian pdf that can be fully determined, in theory at least, from the hypotheses through the relations (3.9) and (3.11). This, of course, precludes from any observational information. For generality, this estimate will be referred to as the **background** state and will be denoted x_B without reference to the time. Correlatively, the error covariance matrix will be called the background error covariance matrix and denoted by \mathbf{B} . Hence, at any time in absence of observations (*i.e* before analysis) the truth state of the atmosphere x^t will have the following Gaussian probability density:

$$p(x^t) = \frac{1}{(2\pi)^{n/2} |\mathbf{B}|^{1/2}} \exp \left[-\frac{1}{2} (x^t - x^b)^T \mathbf{B}^{-1} (x^t - x^b) \right] \quad (3.12)$$

Where x_b is the background state vector obtained by integration of the model.

Note 1:

Dropping the time index has two purposes: 1) since the 3-Dimensional analysis problem which is the one we are interested in is static, the time does not play any particular role; 2) in practice the matrix equation (3.9) cannot be solved for matrices with dimension of 10^5 as it generally occurs in NWP. Consequently, all matrices P_i are approximated by a single stationary matrix \mathbf{B} . Although very crude, this approximation is currently done in every Operational Center. Researches on techniques to better approximate the time series of matrices \mathbf{P}_i $i = 1, \dots$ are underway. In particular, methods based on ensemble forecast seem very promising.

Note 2:

In (3.9) and (3.11) we have assumed that the tangent linear approximation was valid. Evensen (1991) showed on a reduced model that such approximation can lead to unrealistic solution when applied to the estimation of the state of the ocean. However, in practice

the effects of such approximation are very difficult to assess for real size problems, since, as it was previously said, the exact forecast error covariance matrix given by the matrix equation (3.9) cannot be solved.

3.2.5 Observational information

Just before the analysis time, the background information x_b , \mathbf{B} represents the best available estimate. At the analysis time, however, a new source of information is made available under the form of observations. Because observations are real and independent of NWP models, it certainly contains information absent from the background. Consequently, we can legitimately expect an improvement of the current estimate by merging together background and observational information. To do so, we need to quantify the information content in the observations and to explicit how they relate to the state vector.

Direct or "conventional" observations consist in measurements of quantities explicitly represented by the state vector. Typical conventional observations consist in wind, temperature or humidity measurements. At the contrary, indirect or "unconventional" observations, like satellite radiances, does not explicitly enter in the state vector representation, but are functional of the vector. Consequently, indirect observations can only be assimilated if the functional can be numerically evaluated. In that case, the corresponding numerical operator is called the observation operator and generally denoted by H and \mathbf{H} when this operator has the particularity to be linear. For direct observations, the observation operator is merely an interpolation operator. For indirect observation, however, the observation operator can be much more complex. For instance, most of the satellite data are non-linear integral measurements of the temperature or humidity present in the atmosphere.

More precisely, if the p -dimensional vector y^o represents the p observations available at the analysis time and since we are making the hypothesis that the observational error can be expressed by an additive white Gaussian perturbation with known mean and covariance matrix; the relation between the observations and the true state of the atmosphere can be mathematically expressed as:

$$y^o = H[x^t] + \epsilon \quad (3.13)$$

where H is the observation operator and ϵ is a p -dimensional random vector with Gaussian

pdf of and known statistical properties:

$$E\{ \epsilon \} = \mathbf{0} \quad (3.14)$$

$$E\{ \epsilon \epsilon^T \} = \mathbf{R} \quad (3.15)$$

We suppose, in addition, that background and observational errors are not correlated:

$$E\{ \eta \epsilon^T \} = E\{ \epsilon \eta^T \} = \mathbf{0} \quad (3.16)$$

Note that, in general, observations are usually sparse, at least compared to the density of the grid on which the state vector is defined. Hence, in most cases p is smaller than n from at least one or two order of magnitude. Note, also, that 3-dimensional data assimilation neglects some of the temporal aspects in the sense that all the assimilated observations are supposed to be valid at a single time: the analysis time. To avoid redundancy and incoherence, a so-called *assimilation time window* is defined. This is the time period over which the observations are collected. The assimilation time window can be much smaller than the cycling time period. For example, if analyses are performed every 12 hours and a 3-hour time window is used, the collection of observations is likely to begin one or two hours before the analysis time. Furthermore, in real-time applications, there is a certain latency between the time observations are recorded and the time the records are available in Meteorological Centers. In that case, the analysis may not be performed exactly at the supposed analysis time but 1 or 2 hours later. As a results, it might be better to shift a little bit the assimilation time window around the analysis time. In our previous example the data collection could began 2 hours before the analysis time and cut off 1 hour after.

To sum-up, we want to estimate the unknown stochastic vector x^t from the known stochastic vectors x^b and y^o , the stochastic relations (3.1) and (3.13), the statistical relation (3.1-2) and (3.14-16) and the assumption of the availability of an optimal estimate with Gaussian pdf at the time origin of the assimilation cycle.

3.3 Maximum a posteriori (MAP) estimation

The MAP and, more generally, any Bayesian estimate relies on the concepts of *prior* and *posterior* information. These concepts are related to the measurement process which

defines the very moment at which the reality is apprehended. So, information will be said either *prior* or *posterior* depending if it relates to an event that occurs before or after the measurement process. These concepts naturally find their expressions in the different probability densities which are the the mathematical way to quantify information.

We will, denote $p_{x^t}(x)$ the probability for the true state of the atmosphere x^t to be equal to the vector x *prior* to measurement and $p_{y^o/x^t}(y/x)$ the conditional probability for observation vector y^o to be equal to y when the true state of the atmosphere x^t is x . Note that both quantities are prior information since they are valid before the measurement process effectively takes place. They do not, therefore, integrate all the available information. A more interesting function is the *posterior* conditional probability density $p_{x^t/y^o}(x/y)$ which quantifies the probability for the true state of the atmosphere x^t to be equal to x when the observations vectors y^o that was effectively measured is the vector y . Obviously, this function includes the observational information, but it also encapsulates the background information as well. This can be seen from the expression of the Bayes theorem (Jazwinsky 1971):

$$p_{x^t/y^o} = \frac{p_{y^o/x^t} * p_{x^t}}{p_{y^o}} \quad (3.17)$$

Where p_{y^o} is the (unconditional) probability density of the observations. Observe that p_{x^t/y^o} is a random function since it is function of the random variable y^o .

When the actual value y of the observation vector y^o is known, the value x which renders the function $x \rightarrow p_{x^t/y^o}(x/y)$ maximal is called the MAP estimate. This estimate can be interpreted as the state of the atmosphere which is the most likely to have produced the observation vector that was actually measured.

To find the expression of the MAP, we need to evaluate the function p_{x^t/y^o} . For that purpose we have on one hand the expression (3.12) of the **prior** probability density of x^t

$$p_{x^t}(x) = \frac{1}{(2\pi)^{n/2} |\mathbf{B}|^{1/2}} \exp \left[-\frac{1}{2} (x - x^b)^T \mathbf{B}^{-1} (x - x^b) \right] \quad (3.18)$$

And, on the other hand, we can use equations (3.14-16) to derive the prior conditional probability density of the observation vector:

$$p_{y^o/x^t}(x/y) = \frac{1}{(2\pi)^{p/2} |\mathbf{R}|^{1/2}} \exp \left[-\frac{1}{2} (H[x] - y)^T \mathbf{R}^{-1} (H[x] - y) \right] \quad (3.19)$$

Replacing these expressions into equation (3.17), we obtain:

$$\begin{aligned}
p_{x^t/y^o}(x/y) &= \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} * \exp \left[-\frac{1}{2}(H[x] - y)^T \mathbf{R}^{-1}(H[x] - y) \right] \\
&* \frac{1}{(2\pi)^{n/2}|\mathbf{B}|^{1/2}} * \exp \left[-\frac{1}{2}(x - x^b)^T \mathbf{B}^{-1}(x - x^b) \right] \\
&* \frac{1}{p_{y^o}}
\end{aligned} \tag{3.20}$$

Dropping the constants:

$$\begin{aligned}
p_{x^t/y^o}(x/y) &\propto \frac{1}{p_{y^o}} * \exp \left[-\frac{1}{2}(H[x] - y)^T \mathbf{R}^{-1}(H[x] - y) + -\frac{1}{2}(x - x^b)^T \mathbf{B}^{-1}(x - x^b) \right] \\
&\propto \frac{1}{p_{y^o}} * \exp [-J]
\end{aligned} \tag{3.21}$$

In expression (3.21) the vector y is measured and known. Therefore, maximizing $p_{x^t/y^o}(x/y)$ with respect to x is equivalent to minimizing the argument of the exponential function alone. Thus, the MAP estimate is the minimum of the following function:

$$\begin{aligned}
J &= \frac{1}{2}(x - x^b)^T \mathbf{B}^{-1}(x - x^b) + \frac{1}{2}(H[x] - y)^T \mathbf{R}^{-1}(H[x] - y^o) \\
&= \underset{J^b+}{\quad} \underset{J^o}{\quad}
\end{aligned} \tag{3.22}$$

Where, to match the notations defined in Ide *et al.* (1997), we have dropped the distinction between the random variable y^o and a realization y ($y = y^o$) in equation (3.21). The function J is often called the cost function.

At the minimum, the derivative of J vanishes, thus MAP estimate x^a satisfies:

$$\nabla J = 0 = \mathbf{B}^{-1}(x^a - x^b) + \mathbf{H}^T \mathbf{R}^{-1}(Hx^a - y^o) \tag{3.23}$$

Where $\mathbf{H} = \frac{\partial H}{\partial x}$ is the tangent linear approximation of the non linear observation operator H . Solving for x^a , we get:

$$x^a = x^b + [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \mathbf{H} \mathbf{R}^{-1}(y^o - H[x^b]) \tag{3.24}$$

4. VARIATIONAL SOLUTION TO THE MAP ESTIMATION PROBLEM

4.1 Variational (iterative) minimization

In real atmospheric applications, the MAP vector x^a given by equation (3.34) is typically of dimension 10^5 and cannot be directly evaluated. This would require the inversion of the huge matrix $(10^5 \times 10^5)$ $[\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]$ which represents a computational burden that far exceeds the capacity of any existing computer. Instead the solution is sought iteratively using a gradient descent. At each iteration $k > 0$ i An approximate solution x^k of the true solution x^a is recurrently sought as:

$$x_{k+1} = x_k - \mathbf{M}_k \nabla_{x_k} J \quad (4.1)$$

Where \mathbf{M}_k is a prescribed matrix. The iteration is usually initialized with the background vector:

$$x_{k=0} = x^b \quad (4.2)$$

it can be shown that for appropriate matrices \mathbf{M}_k , the sequence defined by (4.1) and (4.2) converges to the MAP solution given by (3.24).

Indeed, solving exactly for x^b in equation (3.34) provides:

$$x^b = x^a - \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} [y^o - H x^a] \quad (4.3)$$

Reporting into (4.1) the expression (3.23) of the gradient ∇J evaluated at x_k , we find:

$$\begin{aligned} x_{k+1} &= x_k - \mathbf{M}_k \nabla J(x_k) \\ &= x_k - \mathbf{M}_k \left[\mathbf{B}^{-1}(x_k - x^b) + \mathbf{H}^T \mathbf{R}^{-1}(H x_k - y^o) \right] \end{aligned} \quad (4.4)$$

Reporting the expression (4.3) of x^b into (4.4) provides:

$$\begin{aligned} x_{k+1} &= x_k - \mathbf{M}_k \nabla J(x_k) \\ &= x_k - \mathbf{M}_k \left[\mathbf{B}^{-1}(x_k - x^a + \mathbf{B} \mathbf{H}^T \mathbf{R}^{-1} [y^o - H x^a]) + \mathbf{H}^T \mathbf{R}^{-1}(H x_k - y^o) \right] \\ &= x_k - \mathbf{M}_k \left[\mathbf{B}^{-1}(x_k - x^a) + \mathbf{H}^T \mathbf{R}^{-1}(H x_k - H x^a) \right] \\ &\simeq x_k - \mathbf{M}_k \left[\mathbf{B}^{-1}(x_k - x^a) + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(x_k - x^a) \right] \\ &\simeq x_k - \mathbf{M}_k \left[\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right] (x_k - x^a) \end{aligned} \quad (4.5)$$

Thus, at each step, the difference between the approximate x_k and exact solution x^a verifies:

$$x_{k+1} - x^a \simeq [\mathbf{I} - \mathbf{M}_k [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]] (x_k - x^a) \quad (4.6)$$

where \mathbf{I} is the identity matrix. So, if at each iteration k , the matrix \mathbf{M}_k is chosen such that:

$$\| \mathbf{I} - \mathbf{M}_k [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}] \| < 1 \quad (4.7)$$

Which means that all the eigenvalues are smaller than 1, then we will have:

$$|x_{k+1} - x^a| < \alpha^n |x_{k=0} - x^a| = \alpha^n |x^b - x^a| \quad \text{with } 0 < \alpha < 1 \quad (4.8)$$

and the sequence x_k will converge to x^a .

There are different ways to construct the matrix sequence \mathbf{M}_k that will satisfy the constraint (4.7). They lead to different minimization algorithm, see Zou *et al.* (1993), for a review of the algorithms currently used in atmospheric sciences. Note that for:

$$\mathbf{M}_k = [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \quad (4.9)$$

We have $x^k = x^a$ and the minimum is reached in one single iteration. This matrix corresponds to the Newton algorithm. This results is not surprising, since the sequence \mathbf{M}_k has been precisely introduced to approximate the matrix $[\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1}$ without inversion. If we further derive the expression (3.23) of the gradient of J with respect to x , we find that the Hessian matrix $\nabla^2 \mathbf{J}$ is equal to:

$$\nabla^2 \mathbf{J} = [\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \quad (4.10)$$

Thus, the Hessian matrix is the step matrix of the Newton algorithm. This result provides the basis for an interesting approach to build the sequence of step matrix \mathbf{M}_k . These techniques, referred to as Quasi-Newton methods aims at approximating the Hessian matrix by building second order information using first order information from the observed behavior of the the cost function and its gradient during the mimimization. For more about quasi-Newton methods see Navon and Legler (1987).

Remark:

It has to be noted that the relation (4.5) is only approximate since it is based on a linearization \mathbf{H} of the non-linear observation operator H . Thus, the convergence of the minimization is only guaranteed for linear observation operators or weakly non-linear observation operators when the increment $x_k - x^a$ remains small. For highly non-linear observation operators, however, there is no theoretical results on the convergence of the method.

4.3 Incremental formulation

Because of non-linearities the minimization has to be conducted by small increment. As a result, the final solution x^a is expected to be not very far from the starting point. Assuming, that the background vector x^b is used as first guess of the minimization (equation (4.2) and dropping the k -index for generality; at each iteration, the solution x is defined by its deviation to the background state:

$$\delta x = x^a - x^b \quad (4.11)$$

which is expected to be remain small.

As a result, for reasonably non-linear model, the linear approximation:

$$H[x] = H[x^b + \delta x] \simeq H[x^b] + \mathbf{H} \delta x \quad (4.12)$$

holds. Reporting δx into the definition (3.22) of the cost function and taking into account (4.11), we can approximate the cost function J by a quadratic function J_{inc} :

$$\begin{aligned} J &= \frac{1}{2}(x - x^b)^T \mathbf{B}^{-1}(x - x^b) + \frac{1}{2}(H[x] - y)^T \mathbf{R}^{-1}(H[x] - y^o) \\ &= \frac{1}{2}v^T \mathbf{B}^{-1} \delta x + \frac{1}{2}(H[x^b + \delta x] - y^o)^T \mathbf{R}^{-1}(H[x^b + \delta x] - y^o) \\ &\simeq \frac{1}{2}v^T \mathbf{B}^{-1}v + \frac{1}{2}(H[x^b] + \mathbf{H} \delta x - y^o)^T \mathbf{R}^{-1}(H[x^b] + \mathbf{H} \delta x - y^o) \\ &\simeq \frac{1}{2}\delta x^T \mathbf{B}^{-1} \delta x + \frac{1}{2}(\mathbf{H} \delta x - d)^T \mathbf{R}^{-1}(\mathbf{H} \delta x - d) \\ &= J_{inc} \end{aligned} \quad (4.13)$$

Where, we have introduced the *innovation vector*:

$$d = y^o - H[x^b] \quad (4.14)$$

The interest in using J_{inc} instead of J is that the former is quadratic and quadratic functionals have a unique minimum. This linearisation solution is also satisfying from a statistical point of view: it corresponds to the classical extended Kalman filter when no *a priori* information is specified on the model error. The incremental formulation has also a computational interest. Pushing further the approximation, a simplified linearized version (lower spectral of spatial resolution) of the observation operator can be used in (4.12). Because the observation operator is applied at each iteration, saving CPU time in its evaluation can result in substantial gains in CPU time for the whole minimization procedure. More about the incremental approach can be found in Courtier *et al.* 1994.

The minimum of J_{inc} will be close to the real solution if the tangent operator \mathbf{H} is a good approximation of the real operator H . This will, of course, depend on the nature of the observation, but, for most data, the tangent linear approximation is an acceptable approximation. For highly non-linear data, the minimization can be split into two loops: an inner quadratic loop linearized around the background state and an outer loop consisting in update of the background vector with the results of the inner loop (Courtier *et al.* 1993).

4.4 Preconditioning

It is well known that numerical methods for matrix inversion are very sensitive to the conditioning of the matrix to invert. Mathematically the conditioning is defined as the ratio of the largest by the lowest eigenvalues and in practice large conditioning number tends to slow down or stop the convergence of iterative inversion methods. At the contrary, matrix with conditioning number close to one will be much easier to invert. Physically, the impact of the conditioning number on the cost function can be illustrated by a function with an elongated elliptic shape for large conditioning number compared to a circular cost function for a conditioning number equal to one. It can immediately be seen that for circular cost function the gradient points exactly toward the minimum, which in that simple illustration corresponds to the origin. While, for elongated elliptic cost function, the gradient has to be significantly rotated in order to obtain the direction of the origin. More about preconditioning can be found in Courtier *et al.* 1994 and Courtier and Fischer 1996.

Since the 3D-VAR data assimilation procedure attempts to invert the Hessian matrix

$[\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]$ and the observation operator H is dependant of the observing system, it is hard to analyse the exact conditioning number in a general way. So, we will concentrate on the first term, which depends on the background error covariance matrix \mathbf{B} only. For that purpose, we decompose the matrix into:

$$\mathbf{B} = \sqrt{\mathbf{B}} \sqrt{\mathbf{B}}^T \quad (4.15)$$

Because a covariance matrix is symmetric positive definite, the matrix $\sqrt{\mathbf{B}}$ defined by (4.15) exists and is unique. Similarly, the same decomposition applies for \mathbf{B}^{-1} :

$$\mathbf{B}^{-1} = \sqrt{\mathbf{B}^{-1}} \sqrt{\mathbf{B}^{-1}}^T \quad (4.16)$$

Introducing the new variable:

$$v = \sqrt{\mathbf{B}^{-1}} \delta x \quad (4.17)$$

the incremental cost function J_{inc} defined by (4.3) becomes:

$$J_{inc} = \frac{1}{2} v^T v + \frac{1}{2} (\mathbf{H} \sqrt{\mathbf{B}} v - d)^T \mathbf{R}^{-1} (\mathbf{H} \sqrt{\mathbf{B}} v - d) \quad (4.18)$$

The gradient of J_{inc} is obtained by differentiating equation (4.18) with respect to v :

$$\begin{aligned} \nabla_v J_{inc} &= v + \sqrt{\mathbf{B}}^T \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H} \sqrt{\mathbf{B}} v - d) \\ &= (\mathbf{I} + \sqrt{\mathbf{B}}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \sqrt{\mathbf{B}}) v - \sqrt{\mathbf{B}}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} d \end{aligned} \quad (4.19)$$

And the Hessian:

$$\nabla_v^2 \mathbf{J}_{inc} = \mathbf{I} + \sqrt{\mathbf{B}}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \sqrt{\mathbf{B}} \quad (4.20)$$

This new Hessian matrix is expected to be much better conditioned than the hessian of the original problem defined by equation (4.10).

The preconditioned formulation of the 3D-VAR data assimilation problem has the additional advantage to avoid the inversion of the matrix \mathbf{B} . It has to be remembered that the background matrix has the same size of the state vector (10^5) which makes it difficult to invert when it contains off-diagonal terms, *i.e* when the spatial correlations are accounted for in the statistical model used to represent the forecast error. This requires, however, to start the minimization with the background vector, so that at the beginning $v = 0$ and there is no need to evaluate equation (4.17).

5. FORECAST ERROR ESTIMATION

In 3D-VAR data assimilation, the background error covariance matrix \mathbf{B} is the only way to propagate observational information backward from observation locations into the model domain. In practice, the more entries in the matrix, the more information is carried away from the observations, so it is very important to have a matrix as full as possible, although a assimilation with a full matrix will certainly not be computationally feasible. Correlatively, as it was outlined in section 4.4, \mathbf{B} plays a major role in the conditioning of the problem. Paramount to this problem is the scaling, *i.e* how the two terms J^b and J^o of the cost function J defined by (3.2) compare. For instance, if the background error covariance matrix is overestimated, the term J^b will have a marginal impact on the global cost function. As a result, the minimization will primarily target the term J^o and will overfit the observations. More serious is the situation when the internal scaling between the analysis variables in \mathbf{B} is not consistent. If, for instance, the temperature errors are underestimated, then the minimization will mainly operate on other variables such as wind, humidity and pressure and temperature will not be adjusted. For these reasons, \mathbf{B} plays a crucial role in the data assimilation problem and should be defined with great care.

Going back to the theoretical expression of \mathbf{B} (set $\mathbf{B} = \mathbf{P}_1$ in equation (3.9)), we see that it is given by a matrix equation. However, it is was pointed it out in the closing remarks of section (3.1.5), this equation cannot be solved in practice due to the excessive computations that the size of \mathbf{B} (10^5) would require. For that reason, as it was also mentioned in section (3.1.5), the same generic matrix is used at every analysis time.

Note

Even so the matrix equation (3.9) could be numerically solved, the problem to find adequate initial conditions (\mathbf{P}_0 in equation (3.6)) would remain. Application of the Kalman filter theory, however, shows that there the memory of these initial conditions is lost after a certain time, and the corresponding filter converge to a asymptotic form. In other words, under certain hypotheses of stationarity and stability of the NWP model the forecast error covariance matrix solution of (3.9) tends to an asymptotic matrix. This matrix can be found by solving the asymptotic form of equation (3.9):

$$0 = \mathbf{M}\mathbf{P}\mathbf{M}^T + \mathbf{Q} \quad (5.1)$$

Where \mathbf{M} and \mathbf{Q} are the asymptotic forms, if they exist, of the model tangent linear approximation \mathbf{M}_{t_0, t_i} and the model error covariance matrix \mathbf{Q}_i . From a theoretical point of view, these results partially legitimize the use of a generic error covariance matrix independently of the analysis time. However, this makes no difference in practice, since we are neither able to solve neither equation (3.9) nor (5.1).

In that context, \mathbf{B} appears as the error covariance matrix of an estimator \mathbf{x}^b of stationary form x^t of the time-variable true atmosphere $x^t(t)$:

$$\mathbf{B} = E\{ (x^b - x^t)(x^b - x^t)^T \} \quad (5.2)$$

As it has been done all along these notes, we will assume that this estimator has a Gaussian pdf.

5.1 Error covariance matrix in two dimensions

To gain so more insight of what can be the structure of \mathbf{B} , let examine the problem in two dimension. Let's consider to random scalar variables x and y , with Gaussian joint pdf. Let's call, respectively, \bar{x} , σ_x^2 , \bar{y} , σ_y^2 the mean and variance of x and y . Let's assume, furthermore, that those two variables are correlated with correlation coefficient ρ_{xy} . Then, the pdf has the following expression (Jazwinsky 1971):

$$p(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{x}}{\sigma_x}\right)^2\right] \times \frac{1}{\sqrt{2\pi}\sigma_y(1 - \rho_{xy}^2)} \exp\left[-\frac{1}{2}\left(\frac{(y - \bar{y} - \rho_{xy}\sigma_x/\sigma_y)(x - \bar{x})}{\sqrt{1 - \rho_{xy}^2}\sigma_y}\right)^2\right] \quad (5.3)$$

The error covariance matrix of the random vector $\begin{bmatrix} x \\ y \end{bmatrix}$ is given by (5.2); which in that particular case provides:

$$\begin{aligned} \mathbf{P}_{xy} &= E\left\{ \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \right) \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} \bar{x} \\ \bar{y} \end{bmatrix} \right)^T \right\} \\ &= \begin{pmatrix} E\{(x - \bar{x})(x - \bar{x})\} & E\{(y - \bar{y})(x - \bar{x})\} \\ E\{(x - \bar{x})(y - \bar{y})\} & E\{(y - \bar{y})(y - \bar{y})\} \end{pmatrix} \end{aligned} \quad (5.4)$$

Using in (5.4) the definition of the mathematical expectation $E\{f\}$ for the joint probability density $p(x, y)$:

$$E\{f\} = \int f(x, y) p(x, y) dx dy \quad (5.5)$$

and plugging the expression (5.2) of $p(x, y)$ into (5.5), we obtain:

$$\mathbf{P}_{\mathbf{xy}} = \begin{pmatrix} \sigma_x^2 & \sigma_x \sigma_y \rho_{xy} \\ \sigma_x \sigma_y \rho_{xy} & \sigma_y^2 \end{pmatrix} \quad (5.6)$$

Thus, the covariance matrix is a symmetric matrix ($\rho_{xy} = \rho_{yx}$) with variances on the diagonal and correlations off-diagonal. It can also be shown (see Robert 1992) that $|\rho_{xy}| \leq 1$ and $\mathbf{P}_{\mathbf{xy}}$ is positive definite.

5.2 Background error covariance matrix for 3D-VAR

The two dimensional shape of the error covariance matrix described by (5.6) can easily be generalized to the dimension n . Thus for the 3D-VAR problem, \mathbf{B} is defined by:

$$\mathbf{B} = \begin{pmatrix} (\sigma_x^2)_{i,i} & (\sigma_y \sigma_x \rho_{y,x})_{i,j} \\ (\sigma_x \sigma_y \rho_{x,y})_{i,j} & (\sigma_y^2)_{i,i} \end{pmatrix} \quad \begin{matrix} 1 \leq i \leq n & 1 \leq j \leq n \\ x = u, v, T, q, P & y = u, v, T, q, P \end{matrix} \quad (5.7)$$

Where:

- 1) $(\sigma_x^2)_{i,i}$ $x = u, v, T, q, P$ are the $5 * n$ variance of the forecast error for the each analysis variable x (wind (u, v) , temperature T , humidity q and pressure P) at each grid point i ($1 \leq i \leq n$).
- 2) $(\rho_{xy})_{i,j}$ are the $5 * n * 4 * (n - 1) / 2$ correlation coefficients between the analysis variable x ($x = u, v, T, q, P$) at grid point i ($1 \leq i \leq n$) and the analysis variable y ($y = u, v, T, q, P$) at grid point j ($1 \leq j \leq n$).

Devising a full background error covariance matrix would, therefore, require $5 * n + 5 * n * 4 * (n - 1) / 2 = 5n * (3n - 2)$ coefficients. The classical approach to estimate these coefficients is to use observations as references or truth. Several studies (Hollingsworth and Loennberg 1986, Loennberg and Hollingsworth 1986, have compared forecasts with observations. However, these studies with observational data are limited to the resolution of the observational network which does not permit to resolve the small scales in the

forecasts. The paucity of data is a recurring problem in operational meteorology and as it was summed up by Dee (1995) the available statistical information necessary for specifying all these elements is severely limited. Courtier *et al.* (1997) estimates that there is 10^7 to 10^8 observation per annum which could only allow estimation of several order of magnitude less elements of the background matrix \mathbf{B} in order to obtain stable statistics under the assumption of ergodicity. These numbers are, of course, valid for observations over the globe. At the mesoscale the actual number of available observations is much smaller while the dimension of the state vector is almost the same (10^5). There is therefore no hope to obtain sufficient statistics from observational studies and in any case when this information were available, it could hardly be store on our present computers (*10Gigabytes*).

5.4 Statistical modeling

Some approximations are, therefore, necessary. At first we will drop the terms involving two variables of different nature. These terms will be regenerated later on by the means of additional constraints added as penalty terms in the cost function. This is indeed a convenient way to define these cross-correlation in a physical way. This issue is addressed in section 5.9. Although, such assumption considerably reduces the number of off-diagonal entries in \mathbf{B} which can now be expressed separately for each variable by $n \times n$ dimensional block.

$$\mathbf{B} = \begin{pmatrix} B^u & & & & \\ & B^v & & & \\ & & B^T & & \\ & & & B^q & \\ & & & & B^P \end{pmatrix} \quad (5.8)$$

Where \mathbf{B}^x are the univariate error covariance matrices for the analysis variables $x = u, v, T, q, P$.

Then, the same statistical model is used to model each univariate error covariance matrices \mathbf{B}^x . In such model, variances are usually derived empirically while correlations are prescribed under the form of a function whose values at grid points provide the correlation coefficients. The functions usually contained a few adjustable parameters, so that there is a certain flexibility to enter some real information in the model. Thiébaux (1974) reviews the properties that such correlation functions must satisfies and suggest some forms that agreed well with the observations in term of fit and spectrum. Lorenc (1992) uses the second order regressive functions defined by Thiébaux (1974) with one adjustable parameter. A

general method to build covariance and correlation function for atmospheric applications has been proposed by Gaspari and Cohn (1999). All these approaches are very similar in their principle, once the shape of the correlation function defined one or two parameters are adjusted so as to fit with the observational evidence. This lead to two distincts problems that are separately addressed below.

The NMC method

Parrish and Derber (1992) used a different strategy, their approach is to accumulate difference between forecasts at different ranges and valid at the same time. The assumption behind the method is that differences between these forecasts represent the forecast error. The main benefit is that it can provides global multivariate correlations with full resolution. The method referred for historical reason as *NMC method* can be described as follows. Let $x_{24}^f(t_a)$ and $x_{12}^f(t_a)$ respectively denote a 24h and 12h forecast valid at the analysis time t_a , we can evaluate the difference:

$$\tilde{x}_{12}^{NMCf}(t_a) = x_{24}^f(t_a) - x_{12}^f(t_a) = M_{t_a, t_a-24}[x^a(t_a-24)] - M_{t_a, t_a-12}[x^a(t_a-12)] \quad (5.9)$$

The NMC method claims that $\tilde{x}^{NMCf}(t_a)$ is representative of the forecast error.

Indeed, Rabier *et al.* 1997 has shown that the error covariance matrix associated with the stochastic process defined by (5.9) can be used as surrogate of the real forecast error covariance matrix defined by (3.9) if the model error does not dominate the short range forecast errors and there is a good global data coverage of the domain. A normalization is, however, necessary if the forecasts range difference $24h - 12h = 12h$ in (5.7), is not the actual time span of an assimilation cycle. Rabier *et al.* 1998 applied a correction factor of 0.9 when using forecasts comparisons of 48 and 24 hour in a 6 hour data assimilation cycle. They found that the statistics obtained from the NMC method compare satisfactorily with the results of the studies performed with observational data. Apart from the NCEP and ECMWF centres, the NMC method has been used at Bureau of Meteorology (Steile *et al.* 1995), Meteo France (Desrozier *et al.* and the UK met office (Ingleby *et al.* 1996).

However, if from the NMC method offer an interesting solution to the shortcoming of statistical information, the practical problem of the handling (storage and computational cost) of this matrix remains. Some approximations still have to be made to reduce the

size of $\tilde{x}^{NMCf}(t_a)$. At NCEP, the off-diagonal terms are simply dropped out. Since this matrix is computed in the spectral space, this is equivalent to assume that the correlation functions are isotropic and homogeneous. A less stringent approximation is made at the European Center where the correlation functions are assumed to be horizontally isotropic. This approximation permits to considerably reduce the matrix (Rabier *et al.* 1998). Unfortunately, such approximation cannot be easily transposed to grid point models like MM5 and a special adaptation of the method is needed for mesoscale applications. Our approach, has been to use the NMC method for the estimation of the variances only and to use statistical model for the correlation coefficients. These are two independent problems that are examined separately in the two following sub-sections.

Forecast error variances in the MM5 3D-VAR system

24 minus 12 hour forecast differences as described by (5.7) have been computed from the NCAR MM5 real time system (Bresch 19??) for all the days of year 1997. These differences have been bined and averaged by month, providing monthly mean forecast error:

$$\bar{x}_{12}^{NMCf}(t_a) = \frac{1}{I} \sum_{i=1}^{i=I} \tilde{x}_{12}^{NMCf}(t_i^a) \quad (5.10)$$

Where I are the number of forecast available at 12z in the month under consideration. Assuming the validity of the NMC method, the ergodicity of the stastitcal properties of the atmosphere, and the pseudo-stationarity of the forecaste error, then the variances of the forecast errors are given by:

$$(\sigma^x)_l = (\bar{x}_{12}^{NMCf}(t_i^a))_l^2 \quad (5.10)$$

Where we have used monthly mean. The structure of σ^T for the month of January and July is shown on figure 1 and 2. The definition of the remaining $4(n-1)^2/2$ correlation coefficients $\rho_{k,l}^x$, $k = 1, \dots, n$, $l = k+1, \dots, n$, $x = u, v, T, q$ is based on a statistical model presented in the following section.

Forecast error correlations in the MM5 3D-VAR system

As said, a full background matrix is not desirable in practice. A classical way to reduce the number of entries of \mathbf{B} is to use homogeneous isotropic correlation functions

depending on a tunable cutoff distance (Cohn et Gaspari 1999). Thus, there will be no entry for any pair of points separated by a distance larger than the cutoff length. Here, homogeneous means that the correlation properties are invariant by geographical translation and isotropic means that these properties are independent of the direction (Gaspari and Cohn 1999). However, great care should be exercised at the cutoff distance to avoid the classical Gibbs phenomenon that usually accompanies function truncations. Gibbs phenomenon can either be damped by use of filters such as Hamming or Lanczos windows or avoided using smooth compactly supported correlation function (Gaspari and Cohn 1999).

A different approach consists in modeling correlation by means of digital filters (Lorenz 1991). The idea behind the use of filters is that we don't need in practice the expression of \mathbf{B} itself, but only to model its effect when multiplied to a vector. The most important effects generated by the product of vector with a covariance matrix is smoothing. This comes from the fact that covariance matrix are self-convoluted (Papoulis 1984) and self convolution has a broadening and smoothing effects (Gaspari and Cohn 1999). As a result, we will model the effect of \mathbf{B} by means of a smoothing filter.

The problem is to find the filter that will give the amount of smoothing needed in the least computations, *i.e.* coefficients. Filter synthesis is a general problem of Signal Processing and among the existing methods, the recursive filter approach is certainly one of the most efficient. The recursive technique allows to create infinite impulse response filters in a finite number of operations. A simple example is offered by the filtering operation consisting in averaging over the m previous points:

$$z_l = \frac{x_{l-m} + \dots + x_l}{m} \quad (5.11)$$

Where x_k , $k = 1, \dots$ is the input sequence and z_l , $l > m$ the output sequence. A priori, the evaluation of the n -dimensional vector $[z_l, \dots, z_{l+n}]^T$ requires $1 + 2 + \dots + n = n(n+1)/2$ operations. Now, the recursive filter defined by:

$$z_{l+1} = x_l + \frac{1}{n+1} (x_{l+1} - z_l) \quad (5.12)$$

requires only 3 operations and thus $3 \times n$ operations to evaluate the n dimensional vector $[z_l, \dots, z_{l+n}]^T$.

Similarly, the recursive filter adopted to represent the correlation of the forecast error is the exponential smoother defined by:

$$z_l = \alpha z_{l-1} + (1 - \alpha) x_l \quad l > 0 \quad (5.13)$$

With $z_0 = 0$, where α is a scalar called the filter parameter $0 < \alpha < 1$. The spatial scales of the filter are directly controlled by the scalar $\beta = (1 - \alpha)$. A simple calculation shows that ($x_l = 0, l < 0$):

$$z_l = (1 - \alpha) (x_l + \alpha x_{l-1} + \dots + \alpha^k x_{l-k} + \dots) \quad (5.14)$$

And using the convolution notation:

$$z_l = (F^+ * x)_l = \sigma_k F_k x_{l-k}, \quad \text{with } F_k^+ = (1 - \alpha)\alpha^k \quad (5.15)$$

Where F^+ denotes the kernel, *i.e* the forward filter defined by (5.14), of the convolution.

In practice, this filtering operation is mono-dimensional and repeated in the two horizontal dimensions. However, the one dimensional forward filtering operation produces a bias that can be removed by application of the corresponding adjoint filter to the output of the forward filter given by (5.13):

$$t_l = \alpha t_{l+1} + (1 - \alpha) z_l \quad l > 0 \quad (5.16)$$

And using the convolution notation:

$$t_l = (F^+ * (F^+ * z))_l \quad (5.17)$$

Where F^- denotes the kernel, *i.e* the adjoint filter defined by (5.17), of the convolution.

Note that the adjoint filter is not causal, but the convolution of the forward and adjoint filter is a new filter F that has the property to be symmetric, *i.e* it is defined for $i = -n, \dots, +n$. The coefficients of this filter are given (Hayden and Purser 1995):

$$F_i = \left(\frac{1 - \alpha}{1 + \alpha} \right) \alpha^{|i|} \quad i = -n, \dots, n \quad (5.18)$$

on Figure 3.

Since F and the forecast error should have the same spectrum, it is interesting to look at the frequential response of the filter F . This function is given by:

$$\hat{F}(w) = \sum_{k=-\infty}^{k=+\infty} \left(\frac{1-\alpha}{1+\alpha} \right)^{|k|} \exp i\omega k \quad (5.19)$$

where the summation has been extended from $-\infty$ to $+\infty$ for generality. The summation can be expanded and after several manipulations described in Annex A4, we find:

$$\hat{F}(w) = \frac{1}{1 + \frac{\alpha}{(1-\alpha)^2} 4 \sin^2 \left(\frac{w}{2} \right)} \quad (5.20)$$

If the filtering operation (forward and adjoint) is repeated L times the frequential response of the corresponding filter F^L will be L times the frequency response of the initial filter F

$$\hat{F}(w)^L = \frac{1}{\left(1 + \frac{\alpha}{(1-\alpha)^2} 4 \sin^2 \left(\frac{w}{2} \right) \right)^L} \quad (5.21)$$

which for low frequency can be expanded into (Lorenc 1991):

$$\hat{F}(w)^L \simeq 1 - \frac{1}{1 + \frac{\alpha}{(1-\alpha)^2}} \frac{L}{w^2} \quad (5.22)$$

Lorenc (1991) shows the similarity between $\hat{F}(w)^L$ and the spectrum of a Gaussian shaped filter as $L \rightarrow \infty$. However, Thiebaut (1975) has criticized the use of Gaussian correlation functions for their insufficient power in the small scales.

For that reason, it is preferable to use instead the *Second Order Auto-Regressive* (SOAR) correlation functions defined by:

$$S(r) = \left(1 + \frac{|r|}{s} \right) \exp \left(-\frac{|r|}{s} \right) \quad (5.23)$$

Where r is the distance separating the two points on which the correlation is calculated and s is a constant. The spectrum of the SOAR function is given by:

$$\hat{S}(w) = \frac{4s}{(1 + w^2 s^2)^2} \quad (5.24)$$

Identifying the two spectrums defined by (5.22) and (5.24), we see that for $L = 2$, both spectrum has the same decay in w^{-4} and if, in addition, alpha is defined by:

$$\alpha = 1 + E - \sqrt{E(E+2)} \quad \text{with } E = \frac{L}{4s^2} \quad (5.25)$$

then, the two spectrums have same values at the origin.

Finally, as it was pointed out by Lorenc (1997), a multiplication by \mathbf{B} does not represent a simple physical operation on the field δx : the physical interpretation of the resulting vector depends on the grid. This comes from the definition of the continuous covariance tensor which \mathbf{B} is supposed to represent. The covariance tensor is obtained as an integral over the domain and, thus, its discretized expression \mathbf{B} involves volume elements of the discrete grid. In fact the covariance matrix \mathbf{B} is a mapper from the dual of the model space on to the model space (Lorenc 1992). In contrast, the filter defined by (5.13) and (5.15) is a mapper from the model space on to the model space (*i.e.* it is grid independant). As a results, one or several pass of the filters cannot represent the correlation \mathbf{B} , but it has to be rescaled by the grid volume elements of the grid (Lorenc 1992).

To sum-up, the background covariance matrix \mathbf{B} used in the numerical experiments presented in section 6 is defined by:

$$\mathbf{B} = \mathbf{D} * F * \mathbf{W}^{-1} F^T \mathbf{D}^T \quad (5.26)$$

Where: \mathbf{D} is the diagonal matrix made of the forecast error variances obtained by application of the NMC method,

F is the recursive filter defined by (5.14) and (5.17), with $\alpha = 0.94$

\mathbf{W} is the n dimensional diagonal matrix of the grid volume elements

operations can be omitted.

6. PRACTICAL IMPLEMENTATION

In this section, we describe the MM5 3D-VAR version 1.0 system that has been used to perform the numerical results presented on the web site at <http://www.mmm.ucar.edu/mm5/3dvar/results/sys1.0/sys1.0tests.html>.

6.1 Analysis variables

In NCEP global analyses, the balanced components of the mass and momentum fields are combined into a single variable (Parrish and Derber 1991). This allows the balance between the mass and momentum fields to be implicitly included in the analysis. In addition, the vorticity and divergence instead of the winds itself has a better conditioning

(Purser 1999, personal communication). Although, a balance constraint is certainly necessary for mesoscale models too, it's not clear yet what kind of constraint should be applied. Research work is underway at NCAR (Snyder et al. 1999, Barker 1999). In the present configuration the analysis variables are horizontal wind velocities (u, v) , temperature (T) mixing ratio (q) and non-hydrostatic pressure.

6.2 Re – initialization with global analyses

Mesoscale data assimilation cycle cannot be repeated indefinitely, but a re-initialisation of the background state is necessary in order to correct the effects of the boundary conditions. The frequency of these re-initialisations is, therefore, dependant of the model domain size and resolution and the synoptic situations. In the presented tests, the cycle has been re-initialized every 24 hour with regional or global analyses.

6.3 Analysis cycle time period

In the present system configuration, only conventional observations can be assimilated. Thus, given the small quantity of conventional observations available on a regional domain at non-synoptic time, it is only meaningful to perform analyses at 00z and 12z. Of course, this 12 hour cycle period will have to be reconsidered when the system has the capability to assimilate data from dense observational system such as radar or satellite.

6.4 Assimilation time window and cut – off time

Ideally, only observations valid at the analysis time t_a should be assimilated. However, depending on the time scale of the phenomena resolved by the model it might make sense to extend the assimilation to observations recorded around the analysis time, typically $t_a \pm 2h$ hours, so that more observations can be used. This extended time period around the analysis time along which observations are accepted is called the **assimilation time window**, the **cut-off** time being assimilation the time window upper boundary, *i.e.* the latest measurement time accepted in the system.

In theory, the analysis and forecast cycle can start immediately after the cut-off time, in practice, however, one have to allow a certain amount of time to account for the observation latency, *i.e.* the time necessary to transmit data from the observing network to the place where the analysis is performed. The longer we wait, the more data are likely to be able to

reach the centers, but also the later the following forecast will start, which is not desirable in an operational context. In general, a time period of one hour seems largely sufficient to allow most of the observations collected at the analysis time to reach meteorological centers.

The MM5 3D-VAR system is hardwired to a cut-off of $t_a + 1$ hour while the time window can be selected by the parameter **TIME_WINDOW** (in hour) from the namelist. In our experiments, we had set the time window to 2 hours, thus all observations valid between $t_a - 1$ hours and $t_a + 1$ hour have been used.

Note that the all the observations present in assimilation time window are assimilated indifferently whether they were recorded exactly at the analysis time or not. Although, this technique allows to increase the number of observations to be assimilate; it has, nonetheless, the drawback to limit the time resolution of the system and requires a special treatment of the observations in order to remove multiple data. By multiple data, we mean observations issued from the at different time from the same station. For example, radiosondes are typically reported every 6 or 12 hours, while most of the wind profilers over the US produces hourly measurements. Thus, a 2-hour time window ($[t_a - 1h, t_a + 1h]$) will three conflicting measurements at each wind profiler location which can leads to serious inconsistency if assimilated. Observations must, therefore, be preably filtered in time. The approach used for the MM5 3D-VAR system was to keep the record closest to the analysis time.

6.5 Observations and observation operators *observations*

The present system can assimilate any observations presented as a model variable (wind, temperature, pressure) in WMO Global Telecommunication System (**GTS**) format. These include, but not limited to, SYNOP and SHIP surface observations, TEMP radiosondes and rawinsondes wind, temperature and relative humidity, PILOT wind observations, AIREP aircraft wind and temperature data, SATEM satellite retrieved temperature observations, SATOB single level wind velocity inferred from cloud motions on geostationary images (see Atlas 1997 for details). Because moisture in MM5 is expressed in term of mixing ratio while relative humidity or dew point measurements are available, these quantities are converted into mixing ratio beforehand according to the relations (Rogers

and Yau, 1989):

$$\begin{aligned}
 E_s &= 6.112 * \exp \left[\frac{17.67 * (T - 273.15)}{(T - 273.15 + 243.5)} \right] \\
 Q_s &= 0.622 * \frac{E_s}{(P - E_s)} \\
 R_h &= 100 * \exp \left[5418.1 * \left(\frac{1}{T} - \frac{1}{T_d} \right) \right] \\
 Q_v &= 0.01 * R_h * Q_s
 \end{aligned}
 \tag{5.1}$$

Where T is the temperature in K, T_d is the dew point in K, P is the pressure in hPa, E_s is the vapor pressure in hPa, Q_s is the saturation mixing ratio in kg/kg, R_h is the relative humidity in % and Q_v is the mixing ratio in kg/kg.

Since, these operations are performed before the the assimilation procedure really start. There is no adjoint operation for these transformations. As a result, the moisture analysis increment is onto the mixing model ratio fiel and not on the temperature nor the pressure. The impact of such approximation is, however, very limited compared to the advantage of having only direct observations (*i.e.* observations expressed in terms of model variables) to assimilate

In the presented experiments, the "syn" (synoptic), "upper" (upper-air), "sao" (satellite) GTS observations files valid at 10z, 11z, 12z and 13z are used for the 12z analysis.

observation operator

Observation operators reduce to 3-dimensional interpolation and are fully linear. The horizontal interpolation is always bilinear while the vertical interpolation depends on the variables: linear for temperature and humidity, logarithmic for wind and non-hydrostatic pressure.

Observations are horizontally referenced with respect to, horizontally, their latitude and longitude. It is, therefore, very important that models and observations lay on the same projection. A classical bilinear 4-point interpolation scheme is used. The vertical interpolation used the model geopotential or "reference" reference height. Physically, this corresponds to sigma level heights. Since the height is constant, both the linear and logarithmic vertical interpolations are linear and the corresponding observation operator is linear.

Note that, although the normal vertical coordinate is the geopotential height, the non-hydrostatic pressure can also be used by setting the key **VERTICAL_COORDINATE** to "**PRESSURE**" in the namelist. If the model pressure is used as vertical coordinate, then pressure observations are not assimilated. This option is only sub-optimal, in the sense in the interpolation package that the vertical coordinate is assumed to be a constant; which is correct for the height but not for the pressure. In other words, there is no pressure analysis increment when the pressure is used as vertical coordinate. Again, this approximation permits to keep the 3D-VAR system fully linear.

6.6 Observational covariance matrix

All observations error are assumed to be uncorrelated in space and time, so that the covariance matrix \mathbf{R} is diagonal. The diagonal which consist of the observational error variance, we use prescribed values derived from the NCEP operational Spectral System Interpolation (SSI) described in Parrish and Derber (1992). Early tests performed with these observational synoptic errors have revealed that the temperature and relative humidity errors were over estimated. Consequently, a correcting factor of 0.5 for the temperature and 0.25 for the relative humidity has been applied. The observational errors currently used in the MM5 3D-VAR system are presented in table 1. These values are defined at four pressure levels $P = 1013mb, 1000mb, 500mb$ and $100mb$. Errors at actual observation locations are vertically linearly interpolated from these levels. These errors are then squared and form as the the diagonal of the observational error covariances matrix. Since the observational error is defined in terms of relative humidity while the control variable for moisture is the mixing ratio, the relation (5.1) is applied to the relative humidity error using the observation temperature and pressure defined at that location.

The assumption of uncorelated observations is an *ad hoc* procedure that dramatically simplifies the numerics of data assimilation systems. This is indeed a rather crude approximation since it's certain that data from a same sounding are vertically correlated and there certainly exist some horizontal correlations when soundings are closed to each other. This is, however, a popular approximation currently in effect at the NCEP (Parrish *et al.* 1997) and at the European Center with the exception of geopotential height (Courtier *et al.* 1997).

6.7 Quality control

At the present time observations quality control is reduced to its simplest expression. For each observation y^o , the innovation vector $y^o - Hx^b$ is evaluated and compared to its expected values σ_{y^o} as defined by the diagonal of the observational error covariance matrix \mathbf{R} . If the innovation vector is five times or more greater than its expected values the observations is rejected:

$$if \frac{|y^o - Hx^b|}{\sigma_{y^o}} > 5 \implies \textit{observation rejected} \quad (6.2)$$

Beside this simple gross check, a convective adjustment correction is also applied to soundings in the observations preprocessing.

6.8 Background vector

Classically, the background vector x_b is given by the previous forecast. At the mesoscale, however, boundary conditions have to be specified and provided by an external larger forecasting system. Because these boundary conditions are not exact, they tend to create noise on the boundaries and it is absolutely necessary to periodically reinitialize the system so as to avoid or, at least, to reduce the propagation of this noise into the domain. Consequently, the mesoscale 3D-VAR system is initialized at 00z with the NCEP regional η -analysis, while the 3, 6, 9 and 12 hour forecasts issued from this analysis are used to create the boundary conditions. The MM5 3D-VAR system is initialized externally once per day at 00z, then all subsequent background vectors are provided internally under the form of previous forecast. However, the NCEP regional analysis valid at 12z and the derived forecast valid at 15z, 18z, 21z and 00Z are used to generate the mesoscale boundary conditions for the rest of the day.

The NCEP Eta model (Rogers *et al.* 1995) is run twice daily starting at 00z and 12z with a forecast range extended to 48h and output every 3 hour. The horizontal resolution of the fields used in MM5 is 80km on 38 vertical levels.

6.9 Background covariance matrix

The background covariance matrix has been described in details in section 5. In the numerical experiments we are presenting, the model forecast errors (diagonal of \mathbf{B}) were taken equal to the observational errors defined in section 6.6. The filter was passed 4 times ($\mathbf{NPASS} = 4$ in the namelist), which produces Gaussian shaped correlation functions with

correlation length of 10 model grid points (**NCORR** = 10); which, with a model resolution of $30km$, represents a physical correlation length of approximately $300km$.

6.10 Minimization procedure

The L-BFGS minimizing routine already implemented in the MM5 adjoint modeling system (Zou *et al.* 1997) has been used. This Limited Memory Quasi Newton algorithm is described in Zou *et al.* 1993. In our experiment, we set the memory to 6, which means that the Hessian matrix is estimated from the six most recent gradient vectors.

6.11 Additional penalty term

It is possible to constraint the analysis by adding penalty terms in the cost function defined by (2.12). For example in the SSI at NCEP, a quadratic term penalizing the difference between the analysis and the background divergence tendencies has been added to the cost function. Such constraint requires the result of the analysis to be in balance. Constraints acts, therefore, as a filter and are very useful to remove noise in the analysis which can generate gravity waves in the subsequent forecast. Such noise also exists at the mesoscale, however, at that scale the requirement for the fields to be in balance can be questionable. A selective filter which render the large structure in balance while keeping the fine structures untouched will be more appropriate. These filters are currently under studies (Snyder *et al.* 1999) and will be implemented ulteriorly. There is at present no additional penalty terms in the formulation of the cost function of the current MM5 3D-VAR system.

REFERENCES

- Bartello P. and H. L. Mitchell, 1992: A continuous three-dimensional model of short-range forecast error covariance. *Tellus*, **44A**, 217-235.
- Barker D. M., 1999: A general formulation for 3DVAR control variables. In preparation.
- Cohn S. E., A. Da Silva, J. Guo, M. Sienkiewicz and D. Lamich, 1999: Assessing the effects of data selection with the DAO Physical-Space Statistical Analysis System. *Mon. Wea. Rev.*, **126**, 2913-2926.

- Cohn S. E., 1997: An introduction to estimation theory. *Journal of the Meteorological Society of Japan*, Vol. 75, No 1B, pp. 257-288.
- Courtier, P., 1997: Dual formulation of four dimensional variational assimilation. *Q.J. R. Meteorol. Soc.*, bf 123, 2449-2461.
- Courtier, P., E. Andersson, W. Heckley, J. Pailleux, D. Vasiljevic, M. Hamrud, A. Hollingworth, F. Rabier and M. Fischer, 1998: The ECMWF implementation of the three dimensional variational (3D-Var). Part I: Formulation. *Q.J. R. Meteorol. Soc.*, **123**, 1-26.
- Dudhia J., 1993: A nonhydrostatic version of the Penn state-NCAR mesoscale model: Validation tests and simulation of an Atlantic cyclone and cold front. *Mon. Wea. Rev.*, 121, 1493-1513.
- Fisher M. and P. Courtier, 1995: Estimating the covariance matrices of analysis and forecast error in variational data assimilation. *ECMWF Technical Memo*. 224, available from ECMWF.
- Gaspari G. and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Q.J. R. Meteorol. Soc.*, March 1999.
- Gelb, A., 1980: *Applied Optimal Estimation*. The M.I.T press, Cambridge, MA.
- Hayden C., M. and R. J. Purser, 1995: Recursive filter objective analysis of meteorological fields: application to NESDIS operational processing. *J. Appl. Meteo.*, **34**, 3-15.
- Hollingworth A, and P. Loennberg, 1986: The statistical structure of short range forecast errors as determined from radiosonde data: Part I. The wind field. *Tellus*, 38A, 111-136.
- Jazwinsky, A. H., 1971: *Stochastic process and filtering theory*. Mathematics in Science and Engineering Vol 64, Academic Press, New York, 376p.
- Loennberg P. and A. Hollingworth, 1986: The statistical structure of short range forecast errors as determined from radiosonde data: Part II. the covariance of height and wind error. *Tellus*, 38A, 137-161.

- Lorenc A. C., 1992: Iterative analysis using covariance functions and filters. *Q.J. R. Meteorol. Soc.*, **118**, 569-591.
- Lorenc A. C., 1997: Development of an operational variational scheme. *Journal of the Meteorological Society of Japan*, Vol. 75, No 1B, pp. 339-346.
- Parrish D. and J. Derber, 1992: The National Meteorological Center Spectral Statistical-Interpolation Analysis, *Mon. Wea. Rev.*, 120, 1747-1763.
- Parrish D., J. Derber, J. Purser, W-S Wu, Z-X Pu, 1997: The NCEP Global Analysis System, recent improvement and future plans. *Journal of the Meteorological Society of Japan*, Vol. 75, No 1B, pp. 359-365.
- Phillips, N. A., 1986: The spatial statistics of random geostrophic modes and first-guess errors. *Tellus*, 38A, 314-322.
- Rabier F., A. Mc Nally, E. Anderson, P. Courtier, P. Uden, J. Eyre, A. Hollingworth, F. Bouttier: 1997: The ECMWF implementation of the three dimensional variational (3D-Var). Part II: Structure function. *Q.J. R. Meteorol. Soc.*, **123**, 26-52.
- Rogers, E., D. Deaven, G. DiMego, 1995: The regional analysis system for the operational "early" eta model: original 80-km configuration and recent changes *Weather and Forecasting*, 10(4): 810-825.
- Snyder C. *et al.*, 1998: A data assimilation program for NCAR. Proposal for the USWRP program, NCAR, Boulder, CO.
- Zou X., I. M. Navon, M. Berger, M. Phua, K. H. Paul, T. Schlick and F. X. LeDimet, 1993: Numerical experience with limited memory quasi-Newton and truncated Newton methods. *SIAM J. on Optimization* 3, 582-608.
- Zou X., Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited area mesoscale model. *Mon. Wea. Rev.*, 124, 2859-2882.
- Zou X., F. Vandenberghe, M. Ponca, Y.-H Kuo, 1997: Introduction to the Adjoint Techniques and the MM5 Adjoint Modelling System. NCAR Technical note, NCAR/TN-435-STR, available from NCAR.